Dynamic multi-scale CNN forest learning for automatic cervical cancer segmentation

Bnouni, Nesrine; Rekik, Islem; Rhim, Mohamed Salah; Amara, Najoua Essoukri Ben

Published in:
Machine Learning in Medical Imaging

DOI:
10.1007/978-3-030-00919-9_3

Publication date:
2018

Document Version
Peer reviewed version

Dynamic Multi-Scale CNN Forest Learning for Automatic Cervical Cancer Segmentation

Nesrine Bnouni[1,2], Islem Rekik[2], Mohamed Salah Rhim[3]
and Najoua Essoukri Ben Amara[1]

[1] LATIS- Laboratory of Advanced Technology and Intelligent Systems, ENISo, Sousse University (www.latis-eniso.org)
[2] BASIRA lab, CVIP, School of Science and Engineering, Computing, University of Dundee, UK (www.basira-lab.com)
[3] Department of Gynecology Obstetrics, Faculty of Medicine of Monastir, Tunisia
nesrine.bnouni@gmail.com

Abstract. Deep-learning based labeling methods have gained unprecedented popularity in different computer vision and medical image segmentation tasks. However, to the best of our knowledge, these have not been used for cervical tumor segmentation. More importantly, while the majority of innovative deep-learning works using convolutional neural networks (CNNs) focus on developing more sophisticated and robust architectures (e.g., ResNet, U-Net, GANs), there is very limited work on how to *aggregate* different CNN architectures to improve their *relational learning* at multiple levels of CNN-to-CNN interactions. To address this gap, we introduce a Dynamic Multi-Scale CNN Forest (C^{K+1}DMF), which aims to address three major issues in medical image labeling and ensemble CNN learning: (1) *heterogeneous* distribution of MRI training patches, (2) a *bi-directional* flow of information between two consecutive CNNs as opposed to cascading CNNs —where information passes in a directional way from current to the next CNN in the cascade, and (3) *multiscale* anatomical variability across patients. To solve the first issue, we group training samples into K clusters, then design a forest with $(K + 1)$ trees: a *principal tree* of CNNs trained using all data samples and *subordinate trees*, each trained using a cluster of samples. As for the second and third issues, we design each dynamic multiscale tree (DMT) in the forest such that each node in the tree nests a CNN architecture. Two successive CNN nodes in the tree pass bidirectional contextual maps to progressively improve the learning of their relational non-linear mapping. Besides, as we traverse a path from the root node to a leaf node in the tree, the architecture of each CNN node becomes shallower to take in smaller training patches. Our C^{K+1}DMF significantly ($p<0.05$) outperformed several conventional and ensemble CNN architectures, including conventional CNN (improvement by 10.3%) and CNN-based DMT (improvement by 5%).

1. Introduction

Nearly more than a quarter of a million patients, diagnosed with cervical cancer, die every year [1]. Magnetic Resonance Imaging (MRI) of the pelvis is the most reliable

imaging modality for staging, treatment planning, and following-up the cervical cancer. In particular, clinical staging of cervical cancer is based on the nodal status [2] and the tumor volume. However, this heavily relies on developing accurate segmentation techniques of cervical cancer on MR images that can effectively capture the large variability in the shape, location, and size of the tumor. To the best of our knowledge, a few fully automatic segmentation methods have been applied to diagnose cervical cancer. In this work [3], a Fisher's linear discriminant analysis approach was used for cervical cancer segmentation. Related works on pelvic organ segmentation leveraged machine learning techniques for automated tissue labeling using Computed Tomography (CT) and MRI data. For instance, advanced organ segmentation methods such as male pelvic organs using non-linear classifiers were proposed in [4, 5]. In [4], a new deformable MR method was devised to segment prostate by integrating deep feature learning with sparse patch matching. In [5], displacement regressors were estimated by multi-task Random Forest (RF) for guiding a deformable segmentation model. On the other hand, several studies showed that the aggregation of multiple classifiers improved the performance of a single classifier. In [6], Choi *et al*. combined different classifier approaches such as RF and Gentle AdaBoost learning techniques for improved classification of mammographic lesion. The approach achieved better results compared to each individual classifier model. However, these methods [3-6] are somewhat limited as they require hand-crafted features and may not be robust against varying image appearances. This motivates us to use Deep Learning (DL) method based on Convolutional Neural Networks (CNNs), where features are extracted and learned automatically for robust segmentation.

Recently, CNNs [7] and Fully Convolutional Network (FCN) [8] have gained more popularity in medical image high-resolution synthesis and organ segmentation as they automatically learn features and have defined the state-of-the-art performance in several tasks. For instance, Bahrami *et al.* [7] devised and trained a deep CNN architecture using appearance (intensity) and anatomical (brain-tissue labels) patches to non-linearly map 3T MRI to 7T MRI. In a different work, Alansary *et al.* [8] used CNN to segment the placenta from the structural T2-w MR scans of the whole uterus. To further boost up the performance of a single CNN, other approaches [9-11] chained a set of CNNs for image labeling and reconstruction. Bahrami *et al.* [9] cascaded deep CNN architectures proposed in [7]. In [10], a combination of three ConvNets was used to detect the heart in a different orthogonal plane. In [11], a cascaded architecture that simultaneously exploits local and global contextual features was utilized to segment brain tumor. However, deep learning methods developed in [7-11] mostly leverage a single CNN architecture. To address this limitation, CNN ensemble learning strategies were developed to leverage the strength of multiple CNN architectures. For example, Kamnitsas *et al*. [12] used ensemble of multiple CNN architectures for brain tumor segmentation. In [13], three paralleled CNNs with the same architecture trained following different image preprocessing methods were used for cervical cancer segmentation. Dolz *et al.* [14] also utilized an ensemble of deep CNNs to segment infant brains in MRI images, where 10 CNNs were combined using majority voting to refine the quality of segmentation. Ju and *al.* [15] investigated a wide spectrum of ensemble learning methods, including the super learner, the Bayes optimal classifier, unweighted averaging and majority voting, with deep neural networks for image recognition tasks. Guan *et al.* [16] demonstrated that their ensemble of deep long short-term memory

(LSTM) networks improved the performance of individual LSTM networks for activity recognition.

All previous ensemble DL methods [7-16] developed for image segmentation and other tasks overlooked three major aspects, which if leveraged would further improve the target results: (1) the *heterogeneity* of MRI training patches distribution, (2) *relational learning* at multiple levels of CNN-to-CNN interactions in the ensemble model (absence of *dynamic* learning), and (3) *multiscale* anatomical variability across patients. On the other hand, recently, Amiri *et al.* [17] designed a Dynamic Multiscale Tree (DMT) structure for brain tumor segmentation, where each node embeds a structured RF classifier or a Bayesian Network (BN) classifier. The DMT cascades classifiers along different traversal paths in a binary tree for multi-label brain image segmentation. Ascending and descending feedbacks between each parent node and its children nodes are dynamically communicated and gradually refined to boost the overall performance of the tree. To address the limitations of aforementioned methods [3-17] while leveraging their appealing aspects, we unprecedentedly propose a *Cluster-based Dynamic Multi-scale Forest* (C^{K+1}DMF) for automatic cervical cancer segmentation. The proposed C^{K+1}DMF comprises $(K+1)$ dynamic multiscale trees (DMTs): a *principal tree* trained using all data samples, and *K subordinate trees*, each trained with a more homogeneous cluster of data samples. In particular, each node, in our proposed DMT, nests a CNN architecture. As we go deeper along the tree edges nearing its CNN leaf nodes, the training patch size decreases and the architecture of each visited CNN node becomes shallower. Ultimately, in the testing stage, a testing intensity patch passes through each of these paralleled DMTs, then majority voting is used to fuse labels from all trees in our dynamic forest.

2. Proposed Cluster-based Dynamic Multi-Scale Dynamic Forest

Fig 1. shows the proposed C^{K+1}DMF architecture. Our method comprises several steps. *First*, we cluster the training intensity patches into K clusters using k-means clustering using squared Euclidean distance. *Second*, we design a dynamic forest, where each tree is defined as a DMT [17], and each node in the DMT nests a CNN architecture. We define a principal tree which will be trained using all patches, and K subordinate trees, each trained using patches belonging to each of the clusters. *Next*, the directional flow $F(n_i \rightarrow n_j)$ from node n_i to node n_j in the tree is defined as: $F(n_i \rightarrow n_j) = L(n_i)$, where $L(n_i)$ is the label map outputted by node n_i. This auto-contextual flow enforces spatial consistency between neighboring patches during training. Finally, to generate the target tumor label map, we apply majority voting to the segmentation maps outputted by each DMT in our forest.

2.1 Root node CNN architecture

Fig. 1-B displays the designed multiscale CNN architectures, including the tree root CNN node in green (i.e., CNN1). To train and test our model, we first automatically define a region of interest (ROI) around the tumor in the cervix, and then we extract 2D intensity patches of size 15×15 from each sagittal slice for each training subject.

Convolutional layers of root node CNN. Our architecture is composed of four layers. The first CNN layer aims to learn a feature representation of the input intensity patch $\boldsymbol{x}$ of size 15×15. The activation function of the first layer is defined as $\boldsymbol{a}_1 = max(0, M(\boldsymbol{f}_1, \boldsymbol{x}) + \boldsymbol{b}_1)$, where $\boldsymbol{f}_1$ corresponds to 64 filters with size 7×7, $\boldsymbol{b}_1$ denotes the biases, and $M(.)$ is the mapping convolution function, followed by ReLU function $max(0,.)$. The activation function in the second layer is formulated as $\boldsymbol{a}_2 = max(0, (\boldsymbol{f}_2, \boldsymbol{a}_1) + \boldsymbol{b}_2)$, where $\boldsymbol{f}_2$ denotes the learned 128 filters of size of 5×5 and $\boldsymbol{b}_2$ the corresponding biases. Similarly, the activation function of the third layer is defined as $\boldsymbol{a}_3 = max(0, (\boldsymbol{f}_3, \boldsymbol{a}_2) + \boldsymbol{b}_3)$. We basically convolve the 128 feature maps of the second layer with 256 filters of size 3×3, followed by ReLU. Finally, we estimate the 256 feature maps of the fourth layer with one filter of size of 3×3, followed by ReLU, using $\boldsymbol{a}_4 = max(0, (\boldsymbol{f}_4, \boldsymbol{a}_3) + \boldsymbol{b}_4)$.

Fully connected layer and regression function. The final layer in our architecture is a fully connected layer followed by a regression function to produce the target label value at the center pixel of the input intensity patch $\boldsymbol{x}$. In fact, during training, each intensity patch is mapped to the center voxel value in its corresponding label patch. The final label map is obtained through probability map thresholding. However, a single CNN has a poor invariance scaling, hence the value of cascaded CNNs [9-11].

2.2 Cascaded CNNs

To further boost the performance of a single CNN, we nest it into a cascade of CNNs while leveraging the performance of the previous CNN using an auto-context model. This is achieved through feed-forwarding the outputted segmentation map of each CNN in the cascade as a context information along with the original intensity map to the next CNN as in [9]. By generating the whole image segmentation map after each CNN, we can enforce spatial consistency as neighboring patches coordinate with each other for joint improvement of tumor segmentation using the next CNN (**Supplementary Fig.1**). However, a CNN nested in such a cascaded architecture, will only benefit from the learning of the previous CNN in the chain. To address this limitation, we leverage the binary dynamic multiscale tree architecture proposed in [17] to define a CNN-based DMT, where each node embeds a specific single-scale CNN architecture.

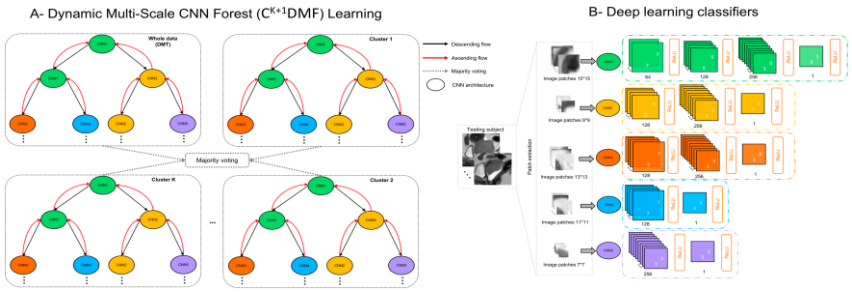

Fig. 1. A) Proposed clustering-based paralleled C^{K+1}DMF for cervical cancer segmentation. We first define a DMT where the training of each tree is dynamic comprising an ascending (red arrow) and descending (black arrow) contextual flow, i.e., probability segmentation map between each parent node and its children nodes. Thus, the two children nodes of the same parent communicate together (via red arrows). Then, we cluster training patches into *K* clusters. Next, we define C^{K+1}DMF, where the principal DMT is trained using the whole training samples, while the subordinate *K* paralleled DMTs are trained using samples in each cluster. Final segmentation map is obtained by applying majority voting to segmentation maps of all C^{k+1}DMF. **B)** Proposed multiscale CNN architectures to define nodes in each tree in our dynamic forest.

2.3 Proposed CNN-based Dynamic Multi-Scale Tree (DMT)

Specifically, our CNN-based DMT is a binary tree $T(V, E)$, where $V = \{n_1, \dots, n_N\}$ denotes the set of N nodes in T and E is a finite set of directed edges connecting two nodes in T. Each parent node nests a CNN architecture, and has two children CNN nodes (one left and one right). At the m^{th} iteration of our DMT learning, a parent node n_i receives the ascending flows $F^m(n_j \to n_i)$ and $F^m(n_{j'} \to n_i)$ from its left and right children nodes n_j and $n_{j'}$, which are first fused then passed on, in a second round $(m+1)$, as contextual information $F^{m+1}(n_i \to n_j) = L^{m+1}(n_i)$ and $F^{m+1}(n_i \to n_{j'}) = L^{m+1}(n_i)$ to the children nodes (**Fig. 2**). We note that each parent CNN node is trained using the fusion of the segmentation maps passed on by its children nodes along with the original input intensity images to produce a new probability map, then communicate it to its two children nodes. Hence, the children CNN nodes of the same parent CNN node cooperate to enhance their parent learning as well as their own learning. This progressively improves the training of each CNN classifier node at each scale level of the tree. Our CNN-DMT is generated automatically as described in **Supplementary Fig.** 3 and **Algorithm 1.**

In addition, to learn how to label anatomical details of the tumor at different scales, we propose a *coarse-to-fine tree learning strategy*, where the root node of the tree embeds the first CNN architecture with a largest patch of size s_1. Next, the architecture of parent node along with its associated training patch size will be copied into its left node (only once as it is not propagated to its grand-child node). The right node, on the other hand, nests a shallower CNN trained with a smaller patch size $s_{l+1} < s_l$ at the next level $l+1$. We note that after copying one CNN architecture once to the left, the left node of the copied CNN will nest a shallower CNN trained at a finer scale. Our proposed CNN-based DMT has two strengths: (1) it captures fine-to-coarse variability in tumor shape and intensity, and (2) both the scale and the patch size influence the performance of parent and children nodes. In fact, to further boost the performance of a single CNN node of the same architecture and to handle unbalanced distribution of negative and positive patches, the left node is copied once to the left. The dynamic tree models the relationships between maps generated from different CNN architecture and at different scales, which helps further reduce the segmentation errors.

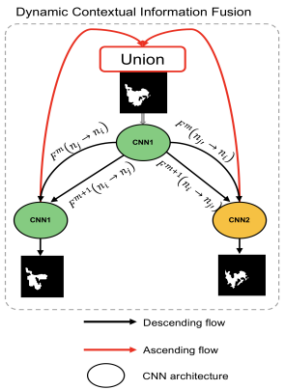

Fig. 2. Binary tree architecture. Ascending and descending flows in our dynamic CNN-DMT.

2.4 Proposed C^{K+1}DMF Learning Framework

Although promising, our CNN-based DMT overlooks the heterogeneity of the training patches. To better explore the heterogeneous intensity patch distribution, we divide all training intensity patches including both positive (inside tumor) and negative (outside tumor) samples into K clusters to define homogeneous patch clusters using k-means

algorithm. Next, we introduce our C^{K+1} Dynamic Multi-scale Forest (DMF), aggregating $(K+1)$ paralleled DMT learners in a unified dynamic architecture. Specifically, our C^{K+1}DMF includes a principal DMT, which is trained using all training patches, and K subordinate DMTs, each trained using patches within a specific cluster. In the testing stage, each testing patch passes through each DMT to eventually estimate the label of its center pixel. Ultimately, we apply majority voting to the segmentation maps outputted by the $(k+1)$ DMTs (**Fig. 1**).

3. Results & Discussion

Dataset, parameters and evaluation. We evaluated our method on 12 clinical T2-weighted MR images of patients with cervical cancer using leave-one-out cross-validation. We use affine registration to align all training MR images to a common space. Each testing subject is then affinely aligned to the training samples. To speed up CNN training, we extract intensity patches within a bounding box of size 141×141 automatically drawn around the cervix and its neighboring regions on sagittal views. For each subject, we extracted 2D intensity patches of size of 15×15 from each sagittal slice for each training subject, thereby producing 363,000 patches for each subject. We used about 4 million patches for training our CNN models. For evaluation, we used the Dice ratio defined as: $Dice = 2A \cap B/A + B$, where A denotes the ground-truth (GT) label map, manually delineated by an expert radiologist, and B the estimated label map.

Evaluation and comparison methods. We compared our C^{K+1}DMF with several segmentation methods (**Table 1**): (1) single CNN [7], (2) Contextual CNN Cascade-1 (CNN-C2), (3) CNN-C3 [9], (4) DMT(d=1), and (5) DMT(d=2) [17]. We note that CNN-C2 cascades two CNNs while CNN-C3 cascades three CNNs. We stop at depth d=2 as the improvement in the classification accuracy becomes negligible. Our C^{k+1}DMF(d=2) is a binary tree with 3 levels (root node + 2 levels). On the other hand, for cluster-based methods, we used the squared Euclidean distance for k-means algorithm. The optimal number of clusters is 3. We note that the number of clusters was empirically set. We also note that DMT(d=1) uses $n = 2^{d+1} - 1$ nodes (i.e., 3 nodes). DMT(d=2) uses 7 nodes. The proposed cluster-based methods include (6)

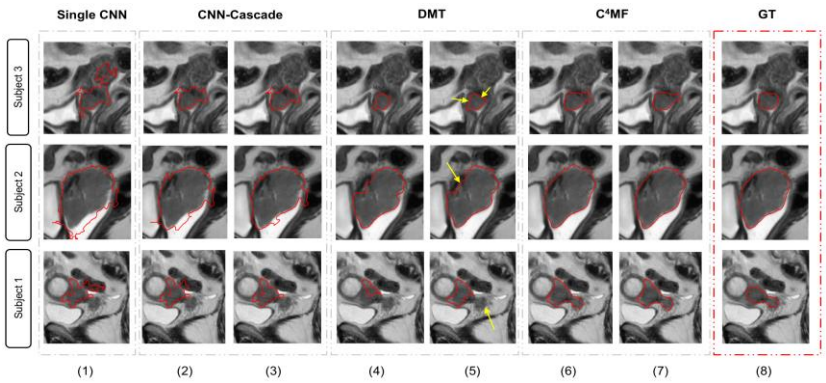

Fig. 3. *Cervical cancer results for 3 representative patients.* Comparison methods and our proposed methods: (1) CNN, (2) CNN-C2, (3) CNN-C3, (4) DMT(1), (5) DMT(2), (6) C^4DMF(1), (7) C^4DMF(2), and (8) GT.

C^{K+1}DMF(d=1), and (7) C^{K+1}DMF(d=2). We note that C^{K+1}DMF(d=1) uses (k+1)×n nodes (i.e., 12 nodes). For instance, C^4DMF (d=2) has 28 nodes.

The detailed experimental results are reported in **Table 1**. We visually inspect the results using our framework and comparison methods for 3 representative patients displayed in **Fig. 3**. Our proposed C^4DMF (d=2) significantly outperformed all comparison methods ($p<0.05$) and produced the best segmentation maps. We note that we report results after applying a post-process using Savitzky-Golay filter for smoothing the tumor boundary outputted by each method. We also note a significant increase in the average Dice ratio from 74.6% by the conventional CNN and 79.9% by the conventional DMT to 84.9% when using C^4DMF(d=2). Our C^4DMF multi-architecture boosted the segmentation accuracy up to 79.9%, in comparison to 76.2% obtained when solely cascaded CNN. Since the subordinate trees in C^4DMF(d=2) are able to better capture the boundary of the tumor, they helped significantly ($p=0.002$) boost up the performance of the principal. This shows that disentangling heterogeneous training samples can help improve CNN training and boost its performance. Both quantitatively and qualitatively, we observed that each cluster captures a unique anatomical detail of the tumor boundary that might be overlooked by other clusters (yellow arrows in **Fig.3** and **Supplementary Fig.2**). A theoretical proof is out of the scope of the current paper.

Table 1. Average Dice ratio using comparison and proposed methods.

Methods	Average Dice (%)
CNN [7]	74.6 ± 4.5
CNN-C2	75.7 ± 6.0
CNN-C3 [9]	76.0 ± 6.1
DMT(d=1)	78.1 ± 7.0
DMT(d=2) [17]	79.9 ± 6.5
C^4-DMF(d=1)	84.0 ± 6.5
C^4-DMF(d=2)	**84.9 ± 6.7**

4. Conclusion

In this paper, we proposed a novel fully automated ensemble deep learning architecture for cervical tumor segmentation on T2-w MR images, leveraging auto-context model to enforce spatial consistency between neighboring patches and introducing paralleled CNN-based dynamic multi-scale trees to handle training-sample heterogeneity. The focus of our paper is not on the choice of the deep learners (CNN or GANs) but on the *boosting* strategy itself by our *generic* clustered forest, which should be applicable to any deep learner (e.g., GANs, UNets, DeepMedic). We could also build our forest using a variety of deep flavors of FCNs (U-Net, GANs). Our boosting strategy is not complex compared to typical boosting algorithms such as ADABOOST. We will extend our C^{K+1}DMF boosting architecture to further handle fine-to-coarse anatomical variability across subjects.

References

1. World Health Organization. (2017). Comprehensive Cervical Cancer Control: A Guide to Essential Practice, 2nd edn. Geneva: WHO; 2014.
2. Bnouni, N., Mechi, O., Rekik, I., Rhim, M. S., & Essoukri Ben Amara, N. (2018, March). Semi-automatic lymph node segmentation and classification using cervical cancer MR

imaging. In *Advanced Technologies for Signal and Image Processing (ATSIP), 2018 4th International Conference on* (pp. 1-6). IEEE.
3. Torheim, T., Malinen, E., Hole, K. H., Lund, K. V., Indahl, U. G., Lyng, H., et al. (2017). Autodelineation of cervical cancers using multiparametric magnetic resonance imaging and machine learning. *Acta Oncologica*, *56*(6), 806-812.
4. Guo, Y., Gao, Y., & Shen, D. (2017). Deformable MR prostate segmentation via deep feature learning and sparse patch matching. In *Deep Learning for Medical Image Analysis* (pp. 197-222).
5. Gao, Y., Shao, Y., Lian, J., Wang, A. Z., Chen, R. C., & Shen, D. (2016). Accurate segmentation of CT male pelvic organs via regression-based deformable models and multi-task random forests. *IEEE transactions on medical imaging*, *35*(6), 1532-1543.
6. Choi, J. Y., Kim, D. H., Plataniotis, K. N., & Ro, Y. M. (2016). Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography. *Expert Systems with Applications*, *46*, 106-121.
7. Bahrami, K., Shi, F., Rekik, I., & Shen, D. (2016). Convolutional neural network for reconstruction of 7T-like images from 3T MRI using appearance and anatomical features. In *Deep Learning and Data Labeling for Medical Applications* (pp. 39-47). Springer, Cham.
8. Alansary, A., Kamnitsas, K., Davidson, A., Khlebnikov, R., Rajchl, M., Malamateniou, C., ... & Kainz, B. (2016, October). Fast fully automatic segmentation of the human placenta from motion corrupted MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 589-597). Springer, Cham.
9. Bahrami, K., Rekik, I., Shi, F., & Shen, D. (2017, September). Joint Reconstruction and Segmentation of 7T-like MR Images from 3T MRI Based on Cascaded Convolutional Neural Networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 764-772). Springer, Cham.
10. Wolterink, J. M., Leiner, T., de Vos, B. D., van Hamersvelt, R. W., Viergever, M. A., & Išgum, I. (2016). Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Medical image analysis*, *34*, 123-136.
11. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., et al. (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, *35*, 18-31.
12. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., et al. (2017). Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. *arXiv preprint arXiv:1711.01468*.
13. Bnouni, N., Ben Amor, H., Rekik, I., Rhim, M. S., Solaiman, B., & Essoukri Ben Amara, N. (2018). Boosting CNN learning by ensemble image preprocessing methods for cervical cancer MR image segmentation, in: International conference on Sensors, Systems, Signals and advanced technologies (SSS).
14. Dolz, J., Desrosiers, C., Wang, L., Yuan, J., Shen, D., & Ayed, I. B. (2017). Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. *arXiv preprint arXiv:1712.05319*.
15. Ju, C., Bibaut, A., & van der Laan, M. J. (2017). The relative performance of ensemble methods with deep convolutional neural networks for image classification. *arXiv preprint arXiv:1704.01664*.
16. Guan, Y., & Plötz, T. (2017). Ensembles of deep lstm learners for activity recognition using wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(2), 11.
17. Amiri, S., Mahjoub, M. A., & Rekik, I. (2018). Tree-based Ensemble Classifier Learning for Automatic Brain Glioma Segmentation. *Neurocomputing, ISSN:0925-2312*.

Dynamic Multi-Scale CNN Forest Learning for Automatic Cervical Cancer Segmentation

SUPPLEMENTARY MATERIAL

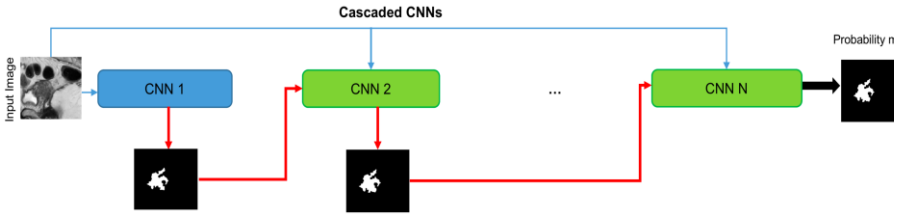

Fig. 3. Proposed contextual CNN cascade for cervical cancer segmentation. We define a CNN cascade where auto-context information, i.e., probability map of the current CNN, is fed-forward to the next CNN (red line), along with the original input image (blue line).

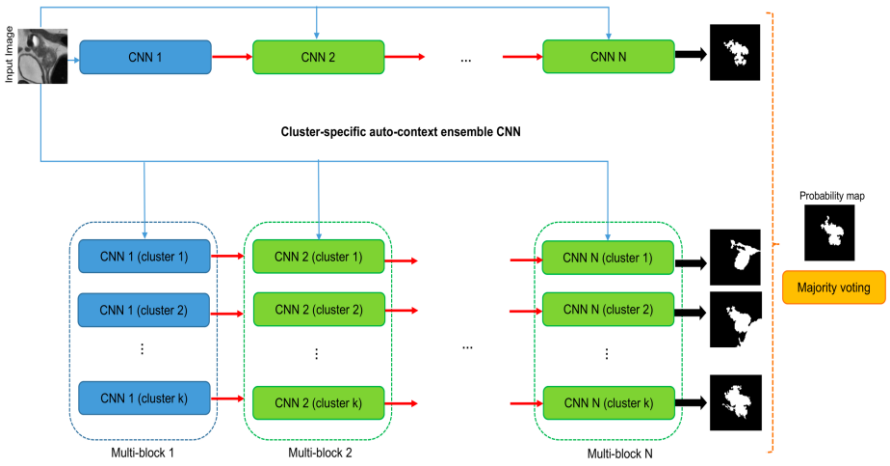

Fig. 2. Proposed cluster-based contextual CNN cascades for cervical cancer segmentation. We first define a CNN cascade where auto-context information, i.e., probability map of the current CNN, is fed-forward to the next CNN (red line), along with the original input image (blue line). Then, we cluster training patches into k clusters. Next, we define $(K + 1)$ CNN cascades, where the first cascade is trained using the whole training samples, while the remaining k paralleled cascades are trained using samples within each cluster. Final segmentation map is obtained by applying majority voting to segmentation maps of all $(K + 1)$ cascades.

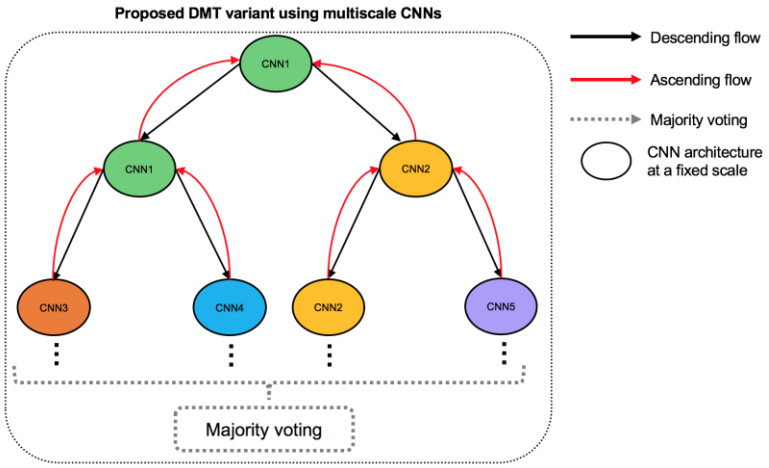

Fig. 3. Proposed CNN-based DMT for cervical cancer segmentation. We define multiple CNN architectures, trained using the whole data, where auto-context information of the parent CNN, is communicated to both its children CNN nodes (black arrow), along with the original input image. In a next round, the auto-context information is communicated from a CNN child to its parent node (red arrow).

Algorithm 1 Proposed CNN-based DMT

Input: Values for number of patches of the first scale for the node right N=1 and for the node left N=2: T(scale=1,N=1), T(scale1,N=2)

Output: Values for number of patches for the Tree T

Procedure:

$scale \leftarrow 1$
while $T(:,:) > 0$ **do**
  $scale \leftarrow scale + 1$
  $j = 1$
  $N \leftarrow 2^{scale}$
  for $i = 1 : N$ **do**
    if i is odd: Node right **then**
      if $j < 2$ **then**
        $T(scale - 1, i) \leftarrow T(scale, round(i/2))$
        $j \leftarrow j + 1$
      else
        $T(scale - 1, i) \leftarrow T(scale, round(i/2)) - 2$
        $j = 1$
      end if
    else {i is even: Node left}
      $T(scale, i) \leftarrow T(scale, i - 1) - 2$
    end if
  end for
end while
return T